

JULY 2009

## Issue Brief

*Dr. Jody Ernst and Richard Wenning*

# Authorizing Matters

### Leave No Charter Behind: An Authorizer Guide to the Use of Growth Data

There is growing consensus in the charter school sector that the success of chartering as an education reform will be measured not by the number of schools in operation (as was the case early on in the movement), but rather the quality of those schools. But what is a “quality school?”<sup>1</sup>

Over the past several years, student growth and growth models have received a lot of attention, in both the charter school sector and the larger education community, as an important indicator for measuring school quality. Educators and policy makers alike have been won over by the argument that because measures of student growth examine the changes in performance of the same students over time, they tell us more about how well schools are educating our nation’s students than do measures of student achievement, or status. Several states now have “federally approved” growth measures as part of their statewide accountability frameworks, but the methods for computing student growth are many and each method tells something different. Understanding exactly what the data is telling is critical if growth models are to have value in evaluating the quality of a given school.

The purpose of this *Issue Brief* is to guide authorizers in the effective use of growth data by outlining the different models for measuring student academic growth and explaining the questions that can be answered by, and the potential limitations of, each type. In addition, this Brief explores key considerations for authorizers as they evaluate their portfolio of schools using a balanced body of evidence, including growth information, and provides examples of authorizers that have effectively incorporated growth measures into their accountability systems.

### Not All Growth Measures Are Alike

The methods for quantifying student progress over time range from the very simple to the very complex, and each method tells us something different. Knowing what questions each method answers (See *Growth Models and the Questions They Address*, page 3) will help authorizers to better utilize the information available, allowing for more informed decisions about the charter schools they oversee. The sections below outline, in plain language, the different types of methods for computing student growth. This Brief is not meant to be a codebook for computing student growth, but rather a helpful resource for effectively using available growth data. (See *State Accountability Frameworks and Growth Models*, page 5, for information on the type of growth models used in each of the 40 states, and the District of Columbia, that have a charter school law).

The methods for quantifying student progress over time range from the very simple to the very complex, and each method tells us something different.



**nacsa**  
NATIONAL ASSOCIATION OF  
CHARTER SCHOOL AUTHORIZERS

**Gain scores:** One of the simplest methods for quantifying student progress is through the computation of gains in scale scores. For students, a gain score is simply this year's score minus last year's score. For a school, student gains are aggregated to answer the question, "What is the difference between this year's average, or median, scale score and last year's average scale score?" While simple to understand and replicate, this approach has many problems associated with it.

Similar to the issues faced by static achievement scores, gains in scale scores are subject to ceiling<sup>2</sup> and cohort<sup>3</sup> effects. In addition, such gains tend to be inversely related to students' initial test scores, so students who scored higher on the initial test tend to have smaller gains than lower scoring students. Gains in scale scores require that the assessments taken by the students be placed on a "vertical scale" allowing one to compare scores from different tests on the same scale. Not all states have such scales. Without them gain score analyses should not be used.

For these reasons, authorizers should be hesitant to use gain scores, especially if other methods of assessing student progress are available. If gain scores are the only measure of progress available, authorizers are cautioned to assign little weight to these values when analyzing a school's quality based on a body of evidence.

**Achievement level transitions:** Other simple methods of assessing student progress are achievement level transitions. These methods address the questions, "What percentage of students have moved up an achievement level (or proficiency band) within the state's accountability system?" or "What percentage of students that were not proficient last year have reached proficiency this year?" There are two commonly used ways of computing values for this method:

- **AYP Growth** which calculates the percentage of students that were not proficient the prior year that are proficient in the current year minus the percentage

of students that dropped below proficiency during that same year.

- **Value Tables** which assigns a numeric value to students that move up (or down) an achievement level.

Data based on achievement level transitions do not recognize growth made by students unless they cross from one achievement level to another and often ignore growth occurring among students that are scoring at the highest levels on state standardized tests. Therefore, unless achievement levels are further subdivided, schools that serve students with profound achievement deficits (e.g. 2 or more years behind in reading and/or math) may not be recognized for helping students progress unless a particular student happens to lie on the cusp of a proficiency category. Similarly, schools are not incentivized to challenge their highest performing students when achievement level transitions are used to assess student progress.

While proficiency is an important goal and reporting progress toward proficiency is good information to have, assessments based on achievement level transitions often fail to provide a great deal more information about the effectiveness of a school than do student scale scores. In fact, the two are rather highly correlated.<sup>4</sup> Therefore, it is not recommended that great weight be given to achievement level transition values when using this data for high-stakes decisions.

Educators and policy makers alike have been won over by the argument that because student growth examines the changes in performance over time, they tell more about how well schools are educating students than do measures of achievement.

**Projection models:** Projection models are more sophisticated and require the use of complicated statistical techniques. These models project a student's future achievement based on prior growth and answer the question, "Is the student projected to reach proficiency (or some other achievement level) within a certain number of years, or by a specific grade?" If a school does not serve students in the grade that the state sets as the goal for proficiency (e.g., all students will be proficient by tenth grade) then alternative goals must be set (for elementary and middle schools, for example). Assuming future growth based on prior growth must still be validated through research. In addition, the projections are only accurate as long as the student continues to show the same amount of growth as

*Authorizing Matters Issue Briefs* are a publication of the National Association of Charter School Authorizers (NACSA), the trusted resource and innovative leader working with public officials and education leaders to increase the number of high quality charter schools in cities and states across the nation. NACSA provides training, consulting, and policy guidance to authorizers and education leaders interested in increasing the number of high quality schools and improving student outcomes.

s/he has made historically. While this allows educators to intervene when students are not projected to reach proficiency, it does not necessarily motivate educators to push students who have already reached proficiency.

Projection measures are best utilized when evaluated in conjunction with a growth measure that specifies each student obtained growth for the academic year, such as a value-added or norm-referenced measure of growth

### **GROWTH MODELS AND THE QUESTIONS THEY ADDRESS**

There are a wide range of questions about student performance that growth data can address. The key to using growth data appropriately and effectively is properly aligning the questions to be asked with the model that can derive such answers.

<b>GROWTH MODEL</b>	<b>QUESTIONS ANSWERED BY THE MODEL</b>
<b>Gain Score</b>	<p>How many points did this student increase, or decrease, between last year and this year?</p> <p>What is the difference between this year's average scores and last year's average scores? Did they increase or decrease?</p> <p>What percentage of the students showed positive gains?</p> <p>Were there between group differences in the percentage of positive gains for ethnic minorities or economically disadvantaged students?</p>
<b>Achievement Level Transitions</b>	<p>What percentage of students moved up an achievement level within the last year? Was this percentage consistent across groups of students?</p> <p>What percentage of students that were not proficient last year have reached proficiency this year? Is this percentage increasing over time?</p>
<b>Projection Models</b>	<p>Is this student projected to reach proficiency (or some other achievement level) within a certain number of years, or by a specific grade?</p> <p>What percentage of students meet or exceed their growth targets, as assessed by the projection model? Does this percentage differ by grade or demographic group?</p>
<b>Value-Added Models</b>	<p>Is school 'A' more effective in raising student achievement than school 'B'?</p> <p>How does student growth in this school compare to what we would expect, relative to other schools, given the characteristics of the student population?</p>
<b>Norm-Referenced Growth</b>	<p>How does this student's growth compare to students with the same achievement history?</p> <p>What is the average growth of students in this group, class, school, or district?</p> <p>Is this school equally effective with students of all ability levels?</p> <p>Is this school becoming more effective over time?</p>
<b>Criterion-Referenced Growth</b>	<p>How much growth is good enough?</p> <p>Was the growth obtained by this student enough to maintain his/her proficiency?</p> <p>Was the growth obtained by this student enough for the student to move toward, or cross into, proficiency?</p>

(described below). To assess a school's effectiveness, it would be necessary to conduct an additional analysis to determine the proportion of students that were successful at meeting, or exceeding, their level of growth projected to occur in the prior year's calculation.

**Value-Added Models:** Value-added models (VAM) are intended to identify and separate out the influence of factors such as students' prior achievement levels, the effect of a given school (or teacher), students' socioeconomic status, and the measurement error associated with the test.<sup>5</sup> This method is very sophisticated and difficult to communicate to lay individuals, but is capable of answering the questions: "Is school 'A' more effective in raising student achievement than school 'B'?" and "How does student growth in a school compare to what we would expect, relative to other schools, given the characteristics of the student population?"

VAM results can provide good information on the effectiveness of charter schools. However, the complexity of the model makes it less likely that authorizers would be able to interpret the outcomes appropriately and use them in an informed way—especially in states like Tennessee and Ohio where the VAMs being used are proprietary and full model content is not disclosed.

**Norm-Referenced Growth:** To date, one normative<sup>6</sup> method for assessing growth has been developed—the use of student growth percentiles (SGP) (Betebenner, 2008). In short, this method assigns a ranking (on a scale from 1-99) for each student based on the students' grade and academic history. At the student level, student growth percentiles address the question, "How does this student's growth compare to students with the same achievement history (i.e., scale score or series of scale scores)?" At the school level, SGPs can be aggregated to answer the question: "How does student growth in my school compare to what we would expect, relative to other schools, given the achievement profile of the student population?" In addition, observed median growth percentiles over time can be used to address the question: "Is my school becoming more effective over time?"

Student growth percentiles allow "apples to apples" comparisons of students and schools, as students of all ability levels are compared only to other similar students, from an academic standpoint. Results using this approach can be described clearly and simply. Descriptions of student growth can be expressed in percentiles to assess progress at the student, school, or system levels. The ability to determine whether growth rates are improving over time, and the ability to relate existing growth percentiles to the growth rates necessary to reach specific standards-based performance levels, make this method a good one for use in high-stakes decision-making.

The one limitation of the growth percentile methodology is the correlation between school size and median growth for a school. It is much easier for small schools to receive extreme (high or low) median growth percentiles in any given year than for larger schools. However, consistently high or low medians over a number of years, even among small schools, provide a good indication of the effectiveness of a school. Therefore, authorizers may want to consider giving a normative growth measure less weight for smaller schools, when making a renewal, or closure, decision.

**Criterion-Referenced Growth:** Criterion-referenced growth is a misnomer of sorts, as it is not a growth calculation in the sense mentioned previously. Criterion-referenced growth qualifies the growth that occurred (using one of several methods) as sufficient or insufficient to get the student up to standard (e.g., catch-up growth) or keep already proficient students in the proficient or advanced range of performance (e.g., keep-up growth).

Criterion-referenced growth is helpful in addressing the question; "How much growth is good enough?" The qualification of growth calculations can be extremely helpful in assessing the effectiveness of a school for students of all ability levels

## Where Does This Leave Authorizers? The Bottom Line on Growth

In states where only gain scores or achievement level transition scores are available, authorizers may consider using a nationally-normed assessment to inform student progress. There are many products available, such as Northwest Evaluation Association's Measure of Academic Progress, Scantron's Performance Series, and the Terra Nova. School operators and authorizers need to research the characteristics of the available products to ensure they are appropriate for the population of students served and weigh the benefits of the assessments for themselves. However, if these methods are not employed in a state's accountability framework, they will not yield direct comparisons to traditional public schools. In any case, the choice of method should be weighed in terms of the questions to be answered, given the methods used state-wide.

Armed with data on student growth and achievement levels (status), authorizers are in a position to consider both when evaluating school performance.

### STATE ACCOUNTABILITY FRAMEWORKS AND GROWTH MODELS<sup>a</sup>

Each state has developed a unique accountability framework for evaluating student and school performance, many of which allow one or more growth analyses. Authorizers should understand the types of growth data available from state assessment systems when considering growth as a key component of charter accountability.

State	Gain Score	Achievement Level Transition	Projection Model	Value-Added	Norm-Referenced	Criterion-Referenced	None
Alaska		■					
Arizona				●			
Arkansas		■					
California	★						
Colorado					■	▼	
Connecticut							◆
Delaware		■					
District of Columbia							◆
Florida <sup>b</sup>	★	■					
Georgia		■					
Hawaii			▲				
Idaho							◆
Illinois							◆
Indiana		■					
Iowa							◆
Kansas		■					
Louisiana		■					
Maryland							◆
Massachusetts <sup>c</sup>		■			■	▼	
Michigan		■	▲				
Minnesota		■					
Mississippi			▲				
Missouri		■					
Nevada							◆
New Hampshire		■					
New Jersey							◆
New Mexico							◆
New York							◆
North Carolina			▲				
Ohio		■	▲	●			
Oklahoma							◆
Oregon							◆
Pennsylvania			▲	●			
Rhode Island							◆
South Carolina							◆
Tennessee		■	▲	●			
Texas <sup>c</sup>			▲				
Utah		■					
Virginia							◆
Wisconsin							◆
Wyoming							◆

a. For the 40 states, and the District of Columbia, which have charter school legislation as of spring 2009.

b. Florida uses a growth-to-standard approach for all students and a value-added growth expectation approach for students in the bottom 25 percent.

c. Massachusetts and Texas adopted new growth models to be implemented in fall of 2009.

USING A BALANCED BODY OF EVIDENCE FOR RENEWAL AND OVERSIGHT: DENVER PUBLIC SCHOOLS

In 2007, Denver Public Schools (DPS) developed a charter school renewal framework that has since been adopted as the school accreditation and rating tool used to annually evaluate all schools in the district, including charter schools.

DPS’s School Performance Framework (or SPF) has been lauded for its comprehensiveness and because it places a greater emphasis on growth in student achievement than on absolute achievement levels.

To determine a school's growth rating, DPS uses seven different growth metrics: the state’s growth rating (High, Typical, or Low), the school’s median growth percentiles (in math, reading, and writing), catch-up growth, keep-up growth, AYP growth, and growth among students that have been continuously enrolled. A school’s rating is also based on student achievement level and achievement gaps (for both minorities and low-income students), as well as other indicators such as college and career readiness, student engagement and satisfaction, demand for seats in the school, and parent and community engagement. In addition, schools are compared to other, demographically similar schools within the district.

The DPS charter school renewal process uses the school’s SPF ratings over the charter term, along with organizational and financial performance ratings, to inform DPS staff recommendations for charter school renewal or closure. Charter schools with a trend of poor academic performance, and who demonstrate a lack of a convincing plan to improve performance, are recommended for closure.

The comprehensive nature of this framework provides schools with transparent expectations and ensures that high stakes decisions are made on multiple points of evidence. Because schools are rated annually, feedback is provided to charters early on. This early warning system allows charters time to make improvements well in advance of contract expiration.

Considering a Balanced Body of Evidence: Growth and Achievement Levels

The focus of this Brief is not an indication that growth is the only data that authorizers should consider when assessing the quality of a school. Many education stakeholders, including federal and state departments of education, consider status a valuable piece of information when evaluating school quality. So, how should authorizers balance growth and status information in judging charter school performance? And whose responsibility is it to collect the data and analyze the results—the school’s, the authorizer’s, or the state’s?

Armed with data on student growth and achievement levels (status), authorizers are in a position to consider both when evaluating school performance. How should authorizers balance the two? While this reflects a policy decision that authorizers may not fully control (given state account-

ability systems), at a minimum authorizers should consider classifying schools in one of four general categories<sup>7</sup> as shown in Figure 1 below.

Using this approach, authorizers can weigh longitudinal growth and achievement levels and distinguish between schools with low average test scores and strong growth (lower right quadrant) and schools with low average test scores and low growth (lower left quadrant). Both categories of schools may not make AYP, but their influence on student progress is different and authorizers should consider them differently when engaging in oversight and renewal decision-making. Likewise, there is a difference between a school with high average test scores that pushes its students further (upper right quadrant) and schools with high scores that do not (upper left quadrant). In building accountability systems that inform high-stakes decision-making, authorizers should carefully consider the balance of growth and achievement data.

FIGURE 1: CATEGORIZING SCHOOLS BY BOTH ACHIEVEMENT LEVEL AND STUDENT LONGITUDINAL GROWTH

Achievement Level-Status (Percent Proficient)	LOW	<----- Longitudinal Growth ----->	HIGH
	Low Achievement Growth Rate & High Achievement Level-Status		High Achievement Growth Rate & High Achievement Level-Status
	Low Achievement Growth Rate & Low Achievement Level-Status		High Achievement Growth Rate & Low Achievement Level-Status



## Who Should Assume Responsibility?

Analyzing growth requires a number of key considerations and careful selection of methodology, particularly when high-stakes decisions, such as renewal or revocation, are based on evaluations of school performance. Authorizers should select a common method for measuring growth among its portfolio of schools and take responsibility for the analysis. Even if a state department of education deploys a method, authorizers should thoroughly understand the limitations of the state approach and the questions it is designed to answer.

In addition, schools may choose their own growth measures to supplement a body of evidence but, for oversight pur-

poses, it's not appropriate to leave this to each school. Assembling longitudinally linked student records across a portfolio of schools and conducting a methodologically sound longitudinal growth analysis is resource-intensive and requires both technical knowledge and technology generally not found within schools<sup>8</sup>.

Moreover, gaining an understanding of the growth rate of students in a given school benefits from the measurement of growth rates of students in other schools, preferably across the entire district or state. For these reasons, it is appropriate and practical for the authorizer to take responsibility for analyzing longitudinal growth across its portfolio of schools.

The Center for Charter Schools at Central Michigan University (CMU) authorizes charter schools within a state that does not yet employ a statewide measure of longitudinal student growth as part of its state accountability system.

Acknowledging the importance of capturing student growth information, in 2002, CMU began requiring the schools it oversees to administer a normative assessment that could be used to measure growth. CMU now requires all elementary and middle schools to administer these assessments each fall and spring. CMU also requires high schools to administer an additional set of assessments that can be administered to students in grades 9, 10 and 11, allowing CMU to capture both growth and status data for high school students that goes beyond the scope of the state's accountability system.

With the consistent use of results from nationally normed assessments, CMU can share data and compare results among the schools in its portfolio. To ensure that teachers and administrators are able to maximize the benefits of these tools, CMU employs two assessment coordinators who provide schools with technical assistance in test administration and professional development in the analysis and use of achievement data. In addition, CMU contracts with expert psychometricians to perform value-added analysis of the growth and status score data. These results are then used to evaluate each school's level of success and, ultimately, help to determine whether CMU will continue to authorize the school.

## Conclusion

This Issue Brief was designed to help authorizers develop strategies to amass a strong body of evidence on the performance of the schools they oversee, and to answer questions critical to judging those schools, such as: "How much learning growth did students at this school experience?" "What was the rate of growth?" "Was the amount or rate of growth good enough?" Each of these questions is important to the oversight of charter schools, but can only be answered by some states' measures of student longitudinal growth.

Charter authorizers should assemble a balanced body of evidence on student and school performance over time and be clear about the questions it does and does not answer. The longitudinal growth of students toward state standards should be a critical part of this body of evidence, and is essential to understanding the productivity of schools. In the absence of a viable state option of assessing student progress, authorizers are encouraged to pursue other ways of collecting and analyzing this information for themselves.

## ENDNOTES

- <sup>1</sup> Agreement on a common definition of quality, one that could cross jurisdiction and state lines, was the focus of the Building Charter School Quality Initiative (BCSQ). With support from the U.S. Department of Education, the National Alliance for Public Charter Schools, the National Association of Charter School Authorizers, the Center for Research on Educational Outcomes, and the Colorado League of Charter Schools, convened a series of meetings of charter school stakeholders, including charter school operators, authorizers, funders, researchers and charter school advocates, from across the country to define charter school quality. The result of this effort is a document entitled, “A Framework for Academic Quality: A Report from the National Consensus Panel on Charter School Academic Quality,” which identifies the core indicators that all stakeholders should use to evaluate school quality. This Issue Brief is based on one of the four key recommendations—Tracking Student Progress Over Time (Growth).
- <sup>2</sup> Ceiling effects result when students scoring at the highest levels of a test have no room left to grow—thus they hit the ceiling.
- <sup>3</sup> Cohort effects result when a group of students from one year look different than a group of students in another year, due to student characteristics, not changes in the learning environment. For example, a third grade class in 2007 may have scored lower, as a group, on the state standardized test than the third grade students in 2006 because a large portion of them entered the third grade not reading at grade level.
- <sup>4</sup> Hill, R. (2006). *Using value tables for a school level accountability system*. National Center for the Improvement of Educational Assessment.
- <sup>5</sup> All tests have measurement error associated with them, including state standardized assessments.
- <sup>6</sup> Normative methods are designed to compare individual students’ test results to those of an appropriate peer group (i.e., norm group) at the state, or national, level; rather than comparing them to some criterion, such as a cut score.
- <sup>7</sup> This multidimensional approach to measuring school performance appears in Doran, H & Izumi, L. (June 2004). *Putting Education to the Test: A Value-Added Model for California*. Pacific Policy Institute.
- <sup>8</sup> For more information about the requirements for conducting longitudinal analysis on a large scale, see Data Quality Campaign. *The 10 Essential Elements in Detail for 2005-06*. [www.dataqualitycampaign.org/activities/elements.cfm#element1](http://www.dataqualitycampaign.org/activities/elements.cfm#element1).

## ABOUT THE AUTHORS

**Dr. Jody Ernst is the Director of Research & Evaluation for the Colorado League of Charter Schools, providing research-based practices for assessing school quality and developing sound performance measures. Her work focuses on developing strategies for assessing schools’ influence on student knowledge gained over time, benchmarking academic growth among at-risk students, and helping organizations set clear and reasonable goals for the students they serve.**

**Richard Wenning is the Associate Commissioner of the Colorado Department of Education, where he is responsible for public policy development and the design and implementation of the state educational accountability system. He previously served as Vice President for Quality and Accountability at the Colorado League of Charter Schools.**



**nacsa**  
NATIONAL ASSOCIATION OF  
CHARTER SCHOOL AUTHORIZERS

National Association of  
Charter School Authorizers  
105 W. Adams Street, Suite 1430  
Chicago, IL 60603-6253

[www.qualitycharters.org](http://www.qualitycharters.org)